

“Shrinkage”

An Article for the Encyclopedia of Environmetrics

George Casella*
Cornell University

September 9, 1999

The problem of estimating the mean of a normal distribution is central to the practice of statistics. This problem is at the heart of many of the most common procedures used today, such as the analysis of variance or regression. If we have a random sample X_1, \dots, X_n , from a normal population with mean μ and variance σ^2 , the natural estimator of μ is the sample mean $\bar{X} = (1/n) \sum_i X_i$. A question of interest is whether this estimator is the best estimator of the parameter μ .

When assessing the performance of an estimator, in particular whether it is best, it is necessary to have a criterion with which to measure it against. A most popular measure is squared error loss, where we measure the performance of an estimator d of a parameter θ by the function

$$(1) \quad L(\theta, d) = (\theta - d)^2,$$

which is called a *loss function*.

Under the loss function (1), \bar{X} has many optimality properties. For example, it is a *minimax estimator* of μ , meaning that of all estimators of μ , its loss has the smallest maximum value. There are other properties that \bar{X} enjoys, including the property of *admissibility*. An estimator d of a parameter θ is an *admissible estimator* of θ under the loss $L(\theta, d)$ if there is no other estimator d' that satisfies

$$E_\theta[L(\theta, d)] \geq E_\theta[L(\theta, d')] \quad \text{for all } \theta,$$

with strict inequality for some values of θ .

*Supported by National Science Foundation Grant DMS-9971586. Email: gc15@cornell.edu. This is paper BU-1451-M in the Department of Biometrics.

Is \bar{X} an admissible estimator of θ ? Hodges and Lehmann (1951) and Blyth (1951) showed that it was. That is, there is no estimator that is uniformly better. However, if the problem is made slightly more complex, an interesting result unfolds. Suppose that instead of estimating the mean of one normal population, we are interested in estimating the mean of many normal populations, that is, we observe \bar{X}_k , $k = 1, \dots, p$, where \bar{X}_k is the mean of n observations from a normal population with mean μ_k and variance σ^2 , and we want to estimate $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)$. The loss of an estimator $\mathbf{d} = (d_1, \dots, d_p)$ is measured by the sum of squared errors, that is

$$(2) \quad L(\boldsymbol{\mu}, \mathbf{d}) = \sum_{k=1}^p (\mu_k - d_k)^2,$$

and we ask if $\bar{\mathbf{X}} = (\bar{X}_1, \dots, \bar{X}_p)$ is still an admissible estimator of $\boldsymbol{\mu}$. If $p = 2$, Stein (1956) showed that the answer is yes, but he also showed that if $p > 2$, the answer is no. Using arguments based on the idea that, for estimating more than 2 means, $\bar{\mathbf{X}}$ tends to be “too long”, Stein demonstrated the existence of a better estimator, a *shrinkage estimator*. Such an estimator shrinks the vector $(\bar{X}_1, \dots, \bar{X}_p)$ toward a specific point in the parameter space. In James and Stein (1961) it was shown that the estimator

$$\mathbf{d}^{JS}(\bar{\mathbf{X}}) = \left(1 - \frac{(p-2)\sigma^2}{|\bar{\mathbf{X}}|^2}\right) \bar{\mathbf{X}}$$

which shrinks $\bar{\mathbf{X}}$ toward 0, uniformly dominates $\bar{\mathbf{X}}$ as an estimator of $\boldsymbol{\mu}$ under the loss (2), so $\bar{\mathbf{X}}$ is not an admissible estimator. This extremely surprising result has resulted in an enormous amount of research in areas such as decision theory and empirical Bayes analysis. Many superior procedures have been since derived. See the review article by Brandwein and Strawderman (1990), or the book by Lehmann and Casella (1998).

References

1. Blyth C. R (1951). On minimax statistical decision procedures and their admissibility. *Ann. Math. Statist.* **22**, 22-42.
2. Brandwein, A. C. and Strawderman, W. E. (1990). Stein estimation: The spherically symmetric case. *Statist. Sci.* **5**, 356-369.

3. Hodges, J. L., Jr., and Lehmann, E. L. (1951). Some applications of the Cramér-Rao inequality. *Proc. Second Berkeley Symp. Math. Statist. Prob.* **1**, University of California Press, 13-22.
4. James, W. and Stein, C. (1961). Estimation with quadratic loss. *Proc. Fourth Berkeley Symp. Math. Statist. Prob.* **1**, University of California Press, 311-319.
5. Lehmann, E. L. and Casella, G. (1998). *Theory of Point Estimation, Second Edition*. New York: Springer-Verlag.
6. Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate distribution. *Proc. Third Berkeley Symp. Math. Statist. Prob.* **1**, University of California Press, 197-206.